

TESIS

**PEMBANGUNAN PERANGKAT LUNAK PERINGKAS
DOKUMEN DARI BANYAK SUMBER BERBASIS WEB
MENGUNAKAN *SENTENCE SCORING* DENGAN METODE
TF-IDF**



FABIANUS HENDY EVAN

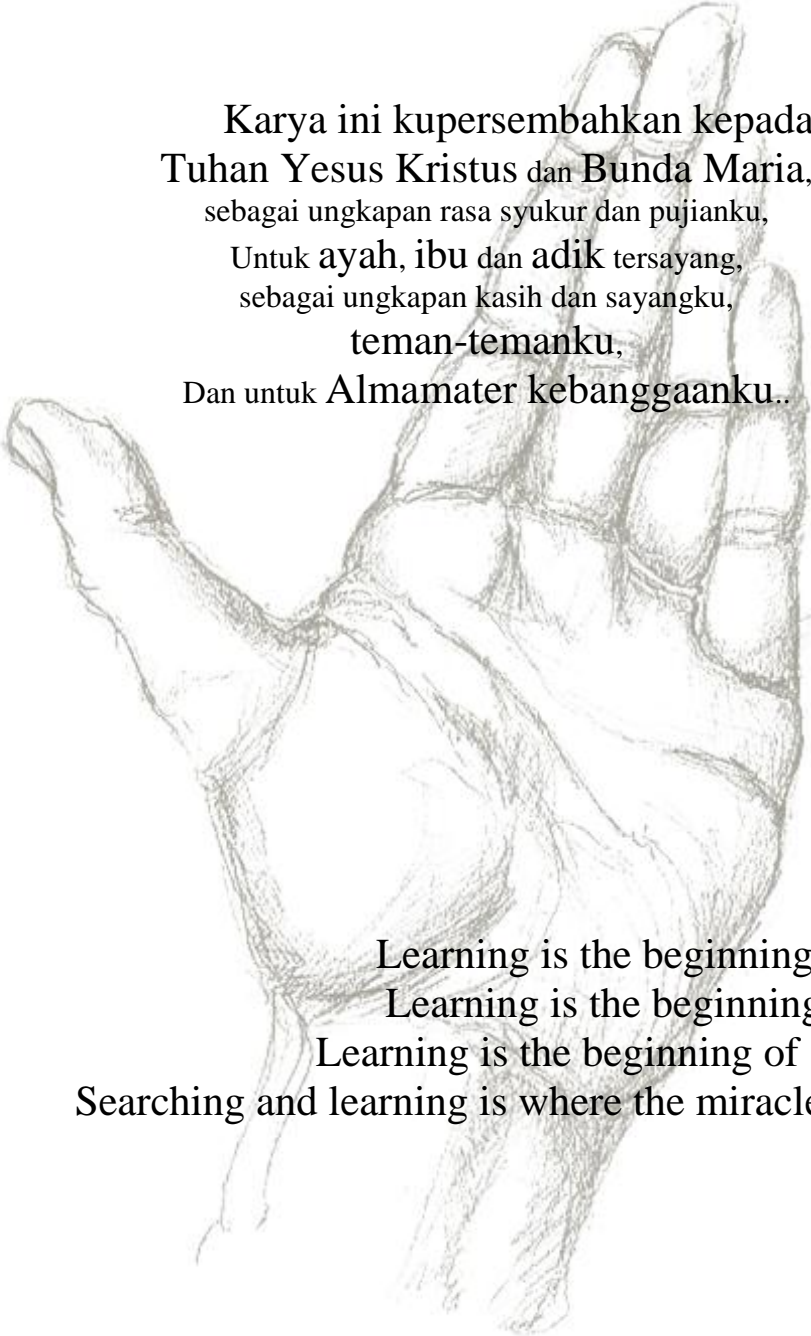
No. Mhs.: 125301915/PS/MTF

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PROGRAM PASCASARJANA

UNIVERSITAS ATMA JAYA YOGYAKARTA

2014



Karya ini kupersembahkan kepada:
Tuhan Yesus Kristus dan Bunda Maria,
sebagai ungkapan rasa syukur dan pujiaku,
Untuk ayah, ibu dan adik tersayang,
sebagai ungkapan kasih dan sayangku,
teman-temanku,
Dan untuk Almamater kebanggaanku..

Learning is the beginning of wealth...
Learning is the beginning of health...
Learning is the beginning of spirituality..
Searching and learning is where the miracle process all
begins...

-Jim Rohn-



UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : FABIANUS HENDY EVAN
Nomor Mahasiswa : 125301915/PS/MTF
Konsentrasi : Mobile Computing
Judul Tesis : Pembangunan Perangkat Lunak Peringkat Dokumen dari
Banyak Sumber berbasis Web menggunakan *Sentence*
Scoring dengan Metode TF-IDF.

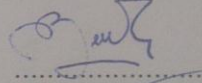
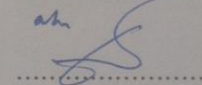
Nama Pembimbing

Tanggal

Tanda Tangan

Dr. Pranowo, S.T., M.T.

10/4/2014


.....

.....

Y. Sigit Purnomo WP., S.T., M.Kom.



UNIVERSITAS ATMA JAYA YOGYAKARTA
PROGRAM PASCASARJANA
PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : FABIANUS HENDY EVAN
Nomor Mahasiswa : 125301915/PS/MTF
Konsentrasi : Mobile Computing
Judul Tesis : Pembangunan Perangkat Lunak Peringkat Dokumen dari
Banyak Sumber berbasis Web menggunakan *Sentence*
Scoring dengan Metode TF-IDF.

Nama Penguji

Dr. Pranowo, S.T., M.T.
(Ketua)

Tanggal

24/4/2014

Tanda Tangan

Y. Sigit Purnomo WP., S.T., M.Kom.
(Anggota)

24/4/2014

Kusworo Anindito, S.T., M.T.
(Anggota)

25/4/2014



Ketua Program Studi

Prof. Ir. Suyoto, M.Sc., Ph.D.

PERNYATAAN

Dengan ini saya :

Nama : Fabianus Hendy Evan

NIM : 125301915

Menyatakan bahwa sepanjang pengetahuan saya dalam penyusunan Tesis, tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah di tulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 2014

Yang menyatakan,

Fabianus Hendy Evan

INTISARI

Kebutuhan untuk meringkas dokumen sangat penting saat kita memasuki era di mana banyak sekali terdapat informasi yang disajikan. Tesis ini menyajikan aplikasi peringkas dokumen otomatis dari banyak sumber berdasarkan pada identifikasi dan ekstraksi kalimat penting dari dokumen yang menjadi sumber informasi menggunakan sentence scoring dengan metode TF-IDF.

Aplikasi ini dibangun berbasis website agar mempermudah pengguna dalam mengakses aplikasi dimanapun dan kapanpun. Aplikasi ini dibangun menggunakan bahasa pemrograman PHP. Penggunaan PHP ditujukan agar proses sentence scoring dapat dilakukan dengan cukup mudah di server dan juga PHP dapat digunakan untuk membuat web service.

Metode TF-IDF memberikan nilai kepada masing-masing kata sesuai dengan tingkat penggunaannya di seluruh sumber dokumen yang ada. Sentence scoring dilakukan dengan menjumlah nilai dari masing-masing kata sehingga didapatkan nilai akhir dari masing-masing kalimat. Kalimat yang akan digunakan sebagai ringkasan di seleksi sesuai dengan peringkat di dalam masing-masing dokumen. Sentence scoring menggunakan metode TF-IDF dapat menghasilkan ringkasan dokumen yang bersifat ekstraktif yang memudahkan pengguna dalam memahami inti dari sebuah dokumen.

Kata-kata kunci: TF-IDF, peringkas dokumen, sentence scoring.

ABSTRACT

The need to summarize document is very important as we enter an era where there is a lot of information presented. This thesis presents an automatic document summarization from various sources based on the identification and extraction of important sentences from the source document information using sentence scoring with TF-IDF method.

This application was built based website that allows users to access applications anywhere and anytime. This application was built using PHP programming language. The use of PHP intended to sentence scoring proses can be done fairly easy in the server and PHP can also be used to create a web service.

TF-IDF method gives value to each word according to their frequency across documents. Sentence scoring is done by adding up the value of each word to obtain the final value of each sentence. Sentence to be used as a summary are selected according to their rank in each document. Sentence scoring using TF-IDF method generate summaries of documents that are extractive and its allows users to understand the gist of a document.

Keywords: TF-IDF, document summarization, sentence scoring.

KATA PENGANTAR

Puji Syukur ke hadirat Tuhan Yang Maha Esa yang telah memberikan rahmatnya dan karunia-Nya, sehingga penulis dapat menyelesaikan laporan Tesis ini dengan baik. Tujuan dari pembuatan Tesis ini adalah sebagai salah satu syarat untuk mencapai derajat Magister Teknik Informatika dari Program Studi Magister Teknik Informatika Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa selama pembuatan Tugas Akhir ini tidak terlepas dari bantuan berbagai pihak yang telah menyumbangkan pikiran, tenaga, dan bimbingan kepada penulis baik secara langsung maupun tidak langsung. Oleh karena itu pada kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya dan bimbingan-Nya sehingga tugas akhir dan penyusunan laporan ini dapat terselesaikan dengan baik.
2. Bapak Prof. Ir. Suyoto, M.Sc., Ph.D. selaku Ketua Program Studi Magister Teknik Informatika Universitas Atma Jaya Yogyakarta.
3. Bapak Dr. Pranowo, M.T. selaku Dosen Pembimbing I yang telah memberikan waktu, tenaga, pikiran, bantuan, dan dukungan kepada penulis sehingga Tugas Akhir ini dapat diselesaikan.
4. Bapak Yohanes Sigit Purnomo WP., S.T., M.Kom. selaku Dosen Pembimbing II yang telah memberikan kepercayaan, bimbingan, dan masukan yang berarti kepada penulis.

5. Bapak Kusworo Anindito, S.T., M.T. selaku Dosen Penguji yang telah memberikan waktunya untuk menguji Tugas Akhir ini.
6. Bapak Eddy Julianto, S.T., M.T. selaku Dosen Penguji yang telah memberikan waktunya untuk menguji Tugas Akhir ini.
7. Seluruh Dosen Universitas Atma Jaya Yogyakarta yang pernah mengajar dan membimbing penulis selama kuliah di Program Studi Magister Teknik Informatika Universitas Atma Jaya Yogyakarta.
8. Orang tua dan keluarga tercinta, yang memberikan dukungan baik moril maupun materiil kepada penulis untuk selalu berusaha mencapai hasil yang terbaik.
9. Untuk Anom, Edo, Ade dan Adven yang telah memberikan dukungan kepada penulis dalam menyelesaikan Tugas Akhir.

Akhir kata penulis menyadari bahwa dalam mengerjakan karya tulis ini masih ada kekurangannya, untuk itu penulis mengharapkan saran dan kritik yang sifatnya membangun, semoga Tugas Akhir ini bisa bermanfaat khususnya bagi penulis sendiri dan pembaca pada umumnya.

Yogyakarta, 2014

Penulis

DAFTAR ISI

HALAMAN JUDUL	I
PERNYATAAN.....	V
INTISARI	VI
ABSTRACT	VII
KATA PENGANTAR.....	VIII
DAFTAR ISI.....	X
DAFTAR TABEL	XII
DAFTAR GAMBAR.....	XIII
PENDAHULUAN.....	14
1.1 Latar Belakang	14
1.2 Rumusan Masalah	16
1.3 Tujuan	16
1.4 Batasan Masalah.....	17
1.5 Manfaat Penelitian	17
1.6 Metodologi	18
1.7 Sistematika Penulisan Tugas Akhir	19
TINJAUAN PUSTAKA	22
2.1 Tinjauan Pustaka	22
2.2 Landasan Teori.....	26
2.2.1 <i>Document Summarization</i>	26
2.2.2 <i>Text Mining</i>	28
2.2.3 <i>TF-IDF (Term Frequency – Inverse Document Frequency)</i>	29
METODOLOGI PENELITIAN	32
3.1 Observasi.....	32
3.2 Studi Literatur	32
3.3 Kuisioner	33
HASIL PENELITIAN DAN PEMBAHASAN	36
4.1 Analisis Sistem.....	36
4.1.1 Perspektif Produk.....	36
4.2 Jenis Antarmuka.....	39
4.2.1 Antarmuka Pemakai	39
4.2.2 Antarmuka perangkat keras	39
4.3 Fungsi Produk	40

4.4	Karakteristik Pengguna	41
4.5	Asumsi dan Ketergantungan	41
4.6	Diagram Aliran Data (<i>Data Flow Diagram</i>).....	41
4.7	Entity Relationship Diagram (ERD)	44
4.8	Perancangan Sistem	45
4.8.1	Perancangan Arsitekur	45
4.8.2	Perancangan Antarmuka	45
4.9	Implementasi dan Pengujian Perangkat Lunak	50
4.10	Implementasi Perangkat Lunak	50
4.10.1	Implementasi dan Alur Peringkasan Dokumen	51
4.10.2	Implementasi Antarmuka.....	72
4.11	Pengujian Perangkat Lunak.....	77
4.12	Pengujian Pengguna	81
4.13	Kelebihan dan Kekurangan Peringkasan Dokumen dari Banyak Sumber menggunakan <i>Sentence Scoring</i> dengan Metode TF-IDF	87
KESIMPULAN DAN SARAN		88
5.1	Kesimpulan	88
5.2	Saran.....	88
DAFTAR PUSTAKA		90

DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka.....	25
Tabel 4.1 Dokumen 1	53
Tabel 4.2 Dokumen 2.....	54
Tabel 4.3 Filter Kata	55
Tabel 4.4 Pembobotan Kata Dokumen 1	57
Tabel 4.5 Pembobotan Kata Dokumen 2	59
Tabel 4.6 Hasil Perhitungan Skor Kalimat Dokumen 1.....	61
Tabel 4.7 Hasil Perhitungan Skor Kalimat Dokumen 2.....	62
Tabel 4.8 Hasil Penggabungan Ringkasan Seluruh Dokumen.....	65
Tabel 4.9 Hasil Pembobotan Kata menggunakan Metode TF-IDF	67
Tabel 4.10 Hasil Perhitungan Skor Kalimat	70
Tabel 4.11 Hasil Pengujian Antarmuka Add Documents	78
Tabel 4.12 Hasil Pengujian Antarmuka Search Documents	79

DAFTAR GAMBAR

Gambar 2.2.1 <i>Multi Document Summarization</i>	27
Gambar 2.2.2 <i>Text Mining</i>	29
Gambar 3.1 Flow Chart Metodologi Penelitian	35
Gambar 4.1 Arsitektur Sistem.....	37
Gambar 4.2 Flow Chart DOSUM	38
Gambar 4.3 DFD Level 0 DOSUM	42
Gambar 4.4 DFD Level 1 DOSUM	42
Gambar 4.5 DFD Level 2 DOSUM	43
Gambar 4.6 ERD Lifepress	44
Gambar 4.7 Perancangan Arsitektur	45
Gambar 4.8 Rancangan Antarmuka Halaman Utama	46
Gambar 4.9 Rancangan Antarmuka Add Documents	47
Gambar 4.10 Rancangan Antarmuka Halaman Search.....	48
Gambar 4.11 Rancangan Antarmuka Halaman Result	49
Gambar 4.12 Halaman Utama.....	72
Gambar 4.13 Halaman Search Documents	73
Gambar 4.14 Halaman Add Documents	74
Gambar 4.15 Halaman Result	76
Gambar 4.16 Grafik kuisisioner IMK website DOSUM	82
Gambar 4.17 Grafik kuisisioner fungsionalitas website DOSUM	85